

# Statistika

(MD360P03Z, MD360P03U)  
ak. rok 2007/2008

Karel Zvára

karel.zvara@mff.cuni.cz  
<http://www.karlin.mff.cuni.cz/~zvara>

9. října 2007



## charakteristiky variability

- ▶ měří nestejnost (**variabilitu**) hodnot spojité veličiny
- ▶ obecně pro míru variability  $s(x)$

$$s(a + x) = s(x),$$

$$s(b \cdot x) = b \cdot s(x), \quad b > 0$$

- ▶ přičtením stejné konstanty  $a$  (posunutím) se charakteristika variability nezmění (nezávisí na poloze)
- ▶ vynásobení kladnou konstantou znamená, že stejnou konstantou nutno vynásobit charakteristiku variability
- ▶ **rozpětí** [range]  $R = x_{(n)} - x_{(1)}$
- ▶ **kvartilové rozpětí** [quartile range]  $R_Q = Q_3 - Q_1$

## charakteristiky variability

- ▶ měří nestejnost (**variabilitu**) hodnot spojité veličiny
- ▶ obecně pro míru variability  $s(x)$

$$s(a + x) = s(x),$$

$$s(b \cdot x) = b \cdot s(x), \quad b > 0$$

- ▶ přičtením stejné konstanty  $a$  (posunutím) se charakteristika variability nezmění (nezávisí na poloze)
- ▶ vynásobení kladnou konstantou znamená, že stejnou konstantou nutno vynásobit charakteristiku variability
- ▶ **rozpětí** [range]  $R = x_{(n)} - x_{(1)}$
- ▶ **kvartilové rozpětí** [quartile range]  $R_Q = Q_3 - Q_1$

## charakteristiky variability

- ▶ měří nestejnost (**variabilitu**) hodnot spojité veličiny
- ▶ obecně pro míru variability  $s(x)$

$$s(a + x) = s(x),$$

$$s(b \cdot x) = b \cdot s(x), \quad b > 0$$

- ▶ přičtením stejné konstanty  $a$  (posunutím) se charakteristika variability nezmění (nezávisí na poloze)
- ▶ vynásobení kladnou konstantou znamená, že stejnou konstantou nutno vynásobit charakteristiku variability
- ▶ **rozpětí** [range]  $R = x_{(n)} - x_{(1)}$
- ▶ **kvartilové rozpětí** [quartile range]  $R_Q = Q_3 - Q_1$

## charakteristiky variability

- ▶ měří nestejnost (**variabilitu**) hodnot spojité veličiny
- ▶ obecně pro míru variability  $s(x)$

$$s(a + x) = s(x),$$

$$s(b \cdot x) = b \cdot s(x), \quad b > 0$$

- ▶ přičtením stejné konstanty  $a$  (posunutím) se charakteristika variability nezmění (nezávisí na poloze)
- ▶ vynásobením kladnou konstantou znamená, že stejnou konstantou nutno vynásobit charakteristiku variability
- ▶ **rozpětí** [range]  $R = x_{(n)} - x_{(1)}$
- ▶ **kvartilové rozpětí** [quartile range]  $R_Q = Q_3 - Q_1$

## charakteristiky variability

- ▶ měří nestejnost (**variabilitu**) hodnot spojité veličiny
- ▶ obecně pro míru variability  $s(x)$

$$s(a + x) = s(x),$$

$$s(b \cdot x) = b \cdot s(x), \quad b > 0$$

- ▶ přičtením stejné konstanty  $a$  (posunutím) se charakteristika variability nezmění (nezávisí na poloze)
- ▶ vynásobení kladnou konstantou znamená, že stejnou konstantou nutno vynásobit charakteristiku variability
- ▶ **rozpětí** [range]  $R = x_{(n)} - x_{(1)}$
- ▶ **kvartilové rozpětí** [quartile range]  $R_Q = Q_3 - Q_1$

## charakteristiky variability

- ▶ měří nestejnost (**variabilitu**) hodnot spojité veličiny
- ▶ obecně pro míru variability  $s(x)$

$$s(a + x) = s(x),$$

$$s(b \cdot x) = b \cdot s(x), \quad b > 0$$

- ▶ přičtením stejné konstanty  $a$  (posunutím) se charakteristika variability nezmění (nezávisí na poloze)
- ▶ vynásobení kladnou konstantou znamená, že stejnou konstantou nutno vynásobit charakteristiku variability
- ▶ **rozpětí** [range]  $R = x_{(n)} - x_{(1)}$
- ▶ **kvartilové rozpětí** [quartile range]  $R_Q = Q_3 - Q_1$

## rozptyl (variance)

- ▶ (výběrový) **rozptyl** (variance) [variance] [VAR.VÝBĚR][var(x)]  
(nevyhovuje druhému požadavku, místo toho:  $s_{a+b \cdot x}^2 = b^2 \cdot s_x^2$ )

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) \\ &= \frac{1}{n-1} \sum_{j=1}^k n_j (x_j^* - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{j=1}^k n_j x_j^{*2} - n \cdot \bar{x}^2 \right) \end{aligned}$$

- ▶ necht'  $x_1 = 1, x_2 = 3, x_3 = 8$ , pak je  
 $\bar{x} = (1 + 3 + 8)/3 = 12/3 = 4$

$$s_x^2 = \frac{1}{3-1} ((1-4)^2 + (3-4)^2 + (8-4)^2) = \frac{26}{2} = 13 \doteq 3,6^2$$



## rozptyl (variance)

- ▶ (výběrový) **rozptyl** (variance) [variance] [VAR.VÝBĚR][var(x)]  
(nevyhovuje druhému požadavku, místo toho:  $s_{a+b \cdot x}^2 = b^2 \cdot s_x^2$ )

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) \\ &= \frac{1}{n-1} \sum_{j=1}^k n_j (x_j^* - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{j=1}^k n_j x_j^{*2} - n \cdot \bar{x}^2 \right) \end{aligned}$$

- ▶ necht'  $x_1 = 1, x_2 = 3, x_3 = 8$ , pak je  
 $\bar{x} = (1 + 3 + 8)/3 = 12/3 = 4$

$$s_x^2 = \frac{1}{3-1} ((1-4)^2 + (3-4)^2 + (8-4)^2) = \frac{26}{2} = 13 \doteq 3,6^2$$

## rozptyl (variance)

- ▶ (výběrový) **rozptyl** (variance) [variance] [VAR.VÝBĚR][var(x)]  
(nevyhovuje druhému požadavku, místo toho:  $s_{a+b \cdot x}^2 = b^2 \cdot s_x^2$ )

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) \\ &= \frac{1}{n-1} \sum_{j=1}^k n_j (x_j^* - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{j=1}^k n_j x_j^{*2} - n \cdot \bar{x}^2 \right) \end{aligned}$$

- ▶ necht'  $x_1 = 1, x_2 = 3, x_3 = 8$ , pak je  
 $\bar{x} = (1 + 3 + 8)/3 = 12/3 = 4$

$$s_x^2 = \frac{1}{3-1} ((1-4)^2 + (3-4)^2 + (8-4)^2) = \frac{26}{2} = 13 \doteq 3,6^2$$

## rozptyl (variance)

- ▶ (výběrový) **rozptyl** (variance) [variance] [VAR.VÝBĚR][var(x)]  
(nevyhovuje druhému požadavku, místo toho:  $s_{a+b \cdot x}^2 = b^2 \cdot s_x^2$ )

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) \\ &= \frac{1}{n-1} \sum_{j=1}^k n_j (x_j^* - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{j=1}^k n_j x_j^{*2} - n \cdot \bar{x}^2 \right) \end{aligned}$$

- ▶ necht'  $x_1 = 1, x_2 = 3, x_3 = 8$ , pak je  
 $\bar{x} = (1 + 3 + 8)/3 = 12/3 = 4$

$$s_x^2 = \frac{1}{3-1} ((1-4)^2 + (3-4)^2 + (8-4)^2) = \frac{26}{2} = 13 \doteq 3,6^2$$

## rozptyl (variance)

- ▶ (výběrový) **rozptyl** (variance) [variance] [VAR.VÝBĚR][var(x)]  
(nevyhovuje druhému požadavku, místo toho:  $s_{a+b \cdot x}^2 = b^2 \cdot s_x^2$ )

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) \\ &= \frac{1}{n-1} \sum_{j=1}^k n_j (x_j^* - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{j=1}^k n_j x_j^{*2} - n \cdot \bar{x}^2 \right) \end{aligned}$$

- ▶ necht'  $x_1 = 1, x_2 = 3, x_3 = 8$ , pak je  
 $\bar{x} = (1 + 3 + 8)/3 = 12/3 = 4$

$$s_x^2 = \frac{1}{3-1} ((1-4)^2 + (3-4)^2 + (8-4)^2) = \frac{26}{2} = 13 \doteq 3,6^2$$

## směrodatná odchylka

- ▶ rozptyl měří průměrný čtverec vzdálenosti od průměru
- ▶ **směrodatná odchylka** [std. deviation]: odmocnina z rozptylu  
[SMODCH.VÝBĚR][sd(x)]

$$s_x = \sqrt{s_x^2}$$

- ▶ zcela vyhovuje požadavkům na míry variability
- ▶ výhoda směrodatné odchylky:  
stejný fyzikální rozměr jako původní data
- ▶ výběrový rozptyl z *třídních* četností:  
Sheppardova korekce (jsou-li všechny intervaly délky  $h$ ):

$$\text{odečti } \frac{h^2}{12}$$

## směrodatná odchylka

- ▶ rozptyl měří průměrný čtverec vzdálenosti od průměru
- ▶ **směrodatná odchylka** [std. deviation]: odmocnina z rozptylu  
[SMODCH.VÝBĚR][sd(x)]

$$s_x = \sqrt{s_x^2}$$

- ▶ zcela vyhovuje požadavkům na míry variability
- ▶ výhoda směrodatné odchylky:  
stejný fyzikální rozměr jako původní data
- ▶ výběrový rozptyl z *třídních* četností:  
Sheppardova korekce (jsou-li všechny intervaly délky  $h$ ):

odečti  $\frac{h^2}{12}$

## směrodatná odchylka

- ▶ rozptyl měří průměrný čtverec vzdálenosti od průměru
- ▶ **směrodatná odchylka** [std. deviation]: odmocnina z rozptylu  
[SMODCH.VÝBĚR][sd(x)]

$$s_x = \sqrt{s_x^2}$$

- ▶ zcela vyhovuje požadavkům na míry variability
- ▶ výhoda směrodatné odchylky:  
stejný fyzikální rozměr jako původní data
- ▶ výběrový rozptyl z *třídních* četností:  
Sheppardova korekce (jsou-li všechny intervaly délky  $h$ ):

odečti  $\frac{h^2}{12}$

## směrodatná odchylka

- ▶ rozptyl měří průměrný čtverec vzdálenosti od průměru
- ▶ **směrodatná odchylka** [std. deviation]: odmocnina z rozptylu  
[SMODCH.VÝBĚR][sd(x)]

$$s_x = \sqrt{s_x^2}$$

- ▶ zcela vyhovuje požadavkům na míry variability
- ▶ výhoda směrodatné odchylky:  
stejný fyzikální rozměr jako původní data
- ▶ výběrový rozptyl z *třídních* četností:  
Sheppardova korekce (jsou-li všechny intervaly délky  $h$ ):

odečti  $\frac{h^2}{12}$



## směrodatná odchylka

- ▶ rozptyl měří průměrný čtverec vzdálenosti od průměru
- ▶ **směrodatná odchylka** [std. deviation]: odmocnina z rozptylu  
[SMODCH.VÝBĚR][sd(x)]

$$s_x = \sqrt{s_x^2}$$

- ▶ zcela vyhovuje požadavkům na míry variability
- ▶ výhoda směrodatné odchylky:  
stejný fyzikální rozměr jako původní data
- ▶ výběrový rozptyl z *třídních* četností:  
Sheppardova korekce (jsou-li všechny intervaly délky  $h$ ):

odečti  $\frac{h^2}{12}$

## příklad – věk matek

- ▶ rozpětí:  $R = 38 - 18 = 20$
- ▶ kvartilové rozpětí:  $R_Q = 28 - 23 = 5$
- ▶ rozptyl

$$\begin{aligned}s^2 &= \frac{1}{98} \left( (26^2 + 35^2 + \dots + 21^2 + 23^2) - 99 \cdot \left( \frac{2544}{99} \right)^2 \right) \\ &= 16,97 \doteq 4,12^2\end{aligned}$$

- ▶ směrodatná odchylka je 4,12

▶ Var. řada věku matek

## příklad – věk matek

- ▶ rozpětí:  $R = 38 - 18 = 20$
- ▶ kvartilové rozpětí:  $R_Q = 28 - 23 = 5$
- ▶ rozptyl

$$s^2 = \frac{1}{98} \left( (26^2 + 35^2 + \dots + 21^2 + 23^2) - 99 \cdot \left( \frac{2544}{99} \right)^2 \right) \\ = 16,97 \doteq 4,12^2$$

- ▶ směrodatná odchylka je 4,12

▶ Var. řada věku matek

## příklad – věk matek

- ▶ rozpětí:  $R = 38 - 18 = 20$
- ▶ kvartilové rozpětí:  $R_Q = 28 - 23 = 5$
- ▶ rozptyl

$$\begin{aligned}s^2 &= \frac{1}{98} \left( (26^2 + 35^2 + \dots + 21^2 + 23^2) - 99 \cdot \left( \frac{2544}{99} \right)^2 \right) \\ &= 16,97 \doteq 4,12^2\end{aligned}$$

- ▶ směrodatná odchylka je 4,12

▶ Var. řada věku matek

## příklad – věk matek

- ▶ rozpětí:  $R = 38 - 18 = 20$
- ▶ kvartilové rozpětí:  $R_Q = 28 - 23 = 5$
- ▶ rozptyl

$$\begin{aligned}s^2 &= \frac{1}{98} \left( (26^2 + 35^2 + \dots + 21^2 + 23^2) - 99 \cdot \left( \frac{2544}{99} \right)^2 \right) \\ &= 16,97 \doteq 4,12^2\end{aligned}$$

- ▶ směrodatná odchylka je 4,12

▶ Var. řada věku matek

## příklad – věk matek 2

- ▶ pomocí třídnicích četností

$$\begin{aligned}s^2 &= \frac{1}{98} \left( (5 \cdot 19^2 + 27 \cdot 22^2 + \dots + 2 \cdot 37^2) - 99 \cdot \left( \frac{2547}{99} \right)^2 \right) \\ &= 16,36 = (4,05)^2\end{aligned}$$

- ▶ navíc Sheppardova korekce

$$s^2 = 16,36 - \frac{3^2}{12} = (3,95)^2$$

## příklad – věk matek 2

- ▶ pomocí třídnicích četností

$$\begin{aligned}s^2 &= \frac{1}{98} \left( (5 \cdot 19^2 + 27 \cdot 22^2 + \dots + 2 \cdot 37^2) - 99 \cdot \left( \frac{2547}{99} \right)^2 \right) \\ &= 16,36 = (4,05)^2\end{aligned}$$

- ▶ navíc Sheppardova korekce

$$s^2 = 16,36 - \frac{3^2}{12} = (3,95)^2$$

## střední odchylka

- ▶ **střední odchylka** [mean deviation]: průměr odchylek od mediánu (někdy od průměru)  $[\text{mean}(\text{abs}(x-\text{median}(x)))]$

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}|$$

- ▶ **střední diference**: průměr vzájemných vzdáleností všech  $n^2$  dvojic

$$\begin{aligned} \Delta &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \\ &= \frac{2}{n^2} \sum_{j>i} (x_{(j)} - x_{(i)}) \end{aligned}$$



## střední odchylka

- ▶ **střední odchylka** [mean deviation]: průměr odchylek od mediánu (někdy od průměru)  $[\text{mean}(\text{abs}(x - \text{median}(x)))]$

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}|$$

- ▶ **střední diference**: průměr vzájemných vzdáleností všech  $n^2$  dvojic

$$\begin{aligned} \Delta &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \\ &= \frac{2}{n^2} \sum_{j>i} (x_{(j)} - x_{(i)}) \end{aligned}$$

## normované charakteristiky rozptýlenosti

- ▶ dosud zavedené charakteristiky variability závisejí na volbě měřítka (např. délka v m nebo v km)
- ▶ hledáme charakteristiky nezávislé na měřítku, nutně *poměrové* měřítko, *kladné* hodnoty
- ▶ umožní **porovnání** z různých souborů
- ▶ **variační koeficient** [sd(x)/mean(x)]

$$v = \frac{s_x}{\bar{x}}$$

- ▶ **(Giniho) koeficient koncentrace**

$$G = \frac{\Delta}{2\bar{x}} \left( = \frac{2 \sum_{i=1}^n i \cdot x_{(i)}}{n \sum_{i=1}^n x_i} - \frac{n+1}{n} \right)$$

například měří nerovnoměrnost příjmů, velikostí územních jednotek, souvisí s plochou u Lorenzovy křivky

## normované charakteristiky rozptýlenosti

- ▶ dosud zavedené charakteristiky variability závisejí na volbě měřítka (např. délka v m nebo v km)
- ▶ hledáme charakteristiky nezávislé na měřítku, nutně *poměrové* měřítko, *kladné* hodnoty
- ▶ umožní **porovnání** z různých souborů
- ▶ **variační koeficient** [sd(x)/mean(x)]

$$v = \frac{s_x}{\bar{x}}$$

- ▶ **(Giniho) koeficient koncentrace**

$$G = \frac{\Delta}{2\bar{x}} \left( = \frac{2 \sum_{i=1}^n i \cdot x_{(i)}}{n \sum_{i=1}^n x_i} - \frac{n+1}{n} \right)$$

například měří nerovnoměrnost příjmů, velikostí územních jednotek, souvisí s plochou u Lorenzovy křivky

## normované charakteristiky rozptýlenosti

- ▶ dosud zavedené charakteristiky variability závisejí na volbě měřítka (např. délka v m nebo v km)
- ▶ hledáme charakteristiky nezávislé na měřítku, nutně *poměrové* měřítko, *kladné* hodnoty
- ▶ umožní **porovnání** z různých souborů
- ▶ **variační koeficient** [sd(x)/mean(x)]

$$v = \frac{s_x}{\bar{x}}$$

- ▶ **(Giniho) koeficient koncentrace**

$$G = \frac{\Delta}{2\bar{x}} \left( = \frac{2 \sum_{i=1}^n i \cdot x_{(i)}}{n \sum_{i=1}^n x_i} - \frac{n+1}{n} \right)$$

například měří nerovnoměrnost příjmů, velikostí územních jednotek, souvisí s plochou u Lorenzovy křivky

## normované charakteristiky rozptýlenosti

- ▶ dosud zavedené charakteristiky variability závisejí na volbě měřítka (např. délka v m nebo v km)
- ▶ hledáme charakteristiky nezávislé na měřítku, nutně *poměrové* měřítko, *kladné* hodnoty
- ▶ umožní **porovnání** z různých souborů
- ▶ **variační koeficient** [sd(x)/mean(x)]

$$v = \frac{s_x}{\bar{x}}$$

- ▶ (Giniho) koeficient koncentrace

$$G = \frac{\Delta}{2\bar{x}} \left( = \frac{2 \sum_{i=1}^n i \cdot x_{(i)}}{n \sum_{i=1}^n x_i} - \frac{n+1}{n} \right)$$

například měří nerovnoměrnost příjmů, velikostí územních jednotek, souvisí s plochou u Lorenzovy křivky

## normované charakteristiky rozptýlenosti

- ▶ dosud zavedené charakteristiky variability závisejí na volbě měřítka (např. délka v m nebo v km)
- ▶ hledáme charakteristiky nezávislé na měřítku, nutně *poměrové* měřítko, *kladné* hodnoty
- ▶ umožní **porovnání** z různých souborů
- ▶ **variační koeficient** [sd(x)/mean(x)]

$$v = \frac{s_x}{\bar{x}}$$

- ▶ **(Giniho) koeficient koncentrace**

$$G = \frac{\Delta}{2\bar{x}} \left( = \frac{2 \sum_{i=1}^n i \cdot x_{(i)}}{n \sum_{i=1}^n x_i} - \frac{n+1}{n} \right)$$

například měří nerovnoměrnost příjmů, velikostí územních jednotek, souvisí s plochou u Lorenzovy křivky

## z-skór, standardizace

- ▶ variační koeficient  $v$ , Giniho koeficient  $G$  – příklady bezrozměrných veličin (zásluhou průměru ve jmenovateli závisí  $G$  i  $v$  na posunutí!)
- ▶ z-skóry `[STANDARDIZE(x;průměr(x);smodch.výběr(x))]`  
\*`[(x-mean(x))/sd(x)]` nebo `[c(scale(x))]`

$$z_i = \frac{x_i - \bar{x}}{s_x}, \quad i = 1, 2, \dots, n$$

- ▶ dostaneme nulový průměr ( $\bar{z} = 0$ ), jednotkový rozptyl ( $s_z = 1$ )
- ▶ z-skóry jsou bezrozměrné  $\Rightarrow$  umožní hodnotit vlastnosti nezávislé na poloze a variabilitě, např. tvar rozdělení
- ▶  $x_1 = 1, x_2 = 2, x_3 = 3 \Rightarrow \bar{x} = 2, s_x = 1$   
 $z_1 = \frac{1-2}{1} = -1, z_2 = \frac{2-2}{1} = 0, z_3 = \frac{3-2}{1} = 1$

## z-skór, standardizace

- ▶ variační koeficient  $v$ , Giniho koeficient  $G$  – příklady bezrozměrných veličin (zásluhou průměru ve jmenovateli závisí  $G$  i  $v$  na posunutí!)
- ▶ z-skóry  $[STANDARDIZE(x;průměr(x);smodch.výběr(x))]$   
\* $[(x-mean(x))/sd(x)]$  nebo  $[c(scale(x))]$

$$z_i = \frac{x_i - \bar{x}}{s_x}, \quad i = 1, 2, \dots, n$$

- ▶ dostaneme nulový průměr ( $\bar{z} = 0$ ), jednotkový rozptyl ( $s_z = 1$ )
- ▶ z-skóry jsou bezrozměrné  $\Rightarrow$  umožní hodnotit vlastnosti nezávislé na poloze a variabilitě, např. tvar rozdělení
- ▶  $x_1 = 1, x_2 = 2, x_3 = 3 \Rightarrow \bar{x} = 2, s_x = 1$   
 $z_1 = \frac{1-2}{1} = -1, z_2 = \frac{2-2}{1} = 0, z_3 = \frac{3-2}{1} = 1$



## z-skór, standardizace

- ▶ variační koeficient  $v$ , Giniho koeficient  $G$  – příklady bezrozměrných veličin (zásluhou průměru ve jmenovateli závisí  $G$  i  $v$  na posunutí!)
- ▶ z-skóry  $[STANDARDIZE(x;průměr(x);smodch.výběr(x))]$   
\* $[(x-mean(x))/sd(x)]$  nebo  $[c(scale(x))]$

$$z_i = \frac{x_i - \bar{x}}{s_x}, \quad i = 1, 2, \dots, n$$

- ▶ dostaneme nulový průměr ( $\bar{z} = 0$ ), jednotkový rozptyl ( $s_z = 1$ )
- ▶ z-skóry jsou bezrozměrné  $\Rightarrow$  umožní hodnotit vlastnosti nezávislé na poloze a variabilitě, např. tvar rozdělení
- ▶  $x_1 = 1, x_2 = 2, x_3 = 3 \Rightarrow \bar{x} = 2, s_x = 1$   
 $z_1 = \frac{1-2}{1} = -1, z_2 = \frac{2-2}{1} = 0, z_3 = \frac{3-2}{1} = 1$

## z-skór, standardizace

- ▶ variační koeficient  $v$ , Giniho koeficient  $G$  – příklady bezrozměrných veličin (zásluhou průměru ve jmenovateli závisí  $G$  i  $v$  na posunutí!)
- ▶ z-skóry  $[STANDARDIZE(x;průměr(x);smodch.výběr(x))]$   
\* $[(x-mean(x))/sd(x)]$  nebo  $[c(scale(x))]$

$$z_i = \frac{x_i - \bar{x}}{s_x}, \quad i = 1, 2, \dots, n$$

- ▶ dostaneme nulový průměr ( $\bar{z} = 0$ ), jednotkový rozptyl ( $s_z = 1$ )
- ▶ z-skóry jsou bezrozměrné  $\Rightarrow$  umožní hodnotit vlastnosti nezávislé na poloze a variabilitě, např. tvar rozdělení
- ▶  $x_1 = 1, x_2 = 2, x_3 = 3 \Rightarrow \bar{x} = 2, s_x = 1$   
 $z_1 = \frac{1-2}{1} = -1, z_2 = \frac{2-2}{1} = 0, z_3 = \frac{3-2}{1} = 1$

## z-skór, standardizace

- ▶ variační koeficient  $v$ , Giniho koeficient  $G$  – příklady bezrozměrných veličin (zásluhou průměru ve jmenovateli závisí  $G$  i  $v$  na posunutí!)
- ▶ z-skóry  $[STANDARDIZE(x;průměr(x);smodch.výběr(x))]$   
\* $[(x-mean(x))/sd(x)]$  nebo  $[c(scale(x))]$

$$z_i = \frac{x_i - \bar{x}}{s_x}, \quad i = 1, 2, \dots, n$$

- ▶ dostaneme nulový průměr ( $\bar{z} = 0$ ), jednotkový rozptyl ( $s_z = 1$ )
- ▶ z-skóry jsou bezrozměrné  $\Rightarrow$  umožní hodnotit vlastnosti nezávislé na poloze a variabilitě, např. tvar rozdělení
- ▶  $x_1 = 1, x_2 = 2, x_3 = 3 \Rightarrow \bar{x} = 2, s_x = 1$   
 $z_1 = \frac{1-2}{1} = -1, z_2 = \frac{2-2}{1} = 0, z_3 = \frac{3-2}{1} = 1$

## z-skór, standardizace

- ▶ variační koeficient  $v$ , Giniho koeficient  $G$  – příklady bezrozměrných veličin (zásluhou průměru ve jmenovateli závisí  $G$  i  $v$  na posunutí!)
- ▶ z-skóry  $[STANDARDIZE(x;průměr(x);smodch.výběr(x))]$   
\* $[(x-mean(x))/sd(x)]$  nebo  $[c(scale(x))]$

$$z_i = \frac{x_i - \bar{x}}{s_x}, \quad i = 1, 2, \dots, n$$

- ▶ dostaneme nulový průměr ( $\bar{z} = 0$ ), jednotkový rozptyl ( $s_z = 1$ )
- ▶ z-skóry jsou bezrozměrné  $\Rightarrow$  umožní hodnotit vlastnosti nezávislé na poloze a variabilitě, např. tvar rozdělení
- ▶  $x_1 = 1, x_2 = 2, x_3 = 3 \Rightarrow \bar{x} = 2, s_x = 1$   
 $z_1 = \frac{1-2}{1} = -1, z_2 = \frac{2-2}{1} = 0, z_3 = \frac{3-2}{1} = 1$

## z-skór, standardizace

- ▶ variační koeficient  $v$ , Giniho koeficient  $G$  – příklady bezrozměrných veličin (zásluhou průměru ve jmenovateli závisí  $G$  i  $v$  na posunutí!)
- ▶ z-skóry  $[STANDARDIZE(x;průměr(x);smodch.výběr(x))]$   
\* $[(x-mean(x))/sd(x)]$  nebo  $[c(scale(x))]$

$$z_i = \frac{x_i - \bar{x}}{s_x}, \quad i = 1, 2, \dots, n$$

- ▶ dostaneme nulový průměr ( $\bar{z} = 0$ ), jednotkový rozptyl ( $s_z = 1$ )
- ▶ z-skóry jsou bezrozměrné  $\Rightarrow$  umožní hodnotit vlastnosti nezávislé na poloze a variabilitě, např. tvar rozdělení
- ▶  $x_1 = 1, x_2 = 2, x_3 = 3 \Rightarrow \bar{x} = 2, s_x = 1$   
 $z_1 = \frac{1-2}{1} = -1, z_2 = \frac{2-2}{1} = 0, z_3 = \frac{3-2}{1} = 1$

## charakteristiky tvaru: šikmost

- ▶ invariantní vůči posunutí i změně měřítka:

$$\gamma(\mathbf{a} + \mathbf{x}) = \gamma(\mathbf{x})$$

$$\gamma(\mathbf{b} \cdot \mathbf{x}) = \gamma(\mathbf{x}) \quad \mathbf{b} > 0$$

- ▶ šikmost  $\sqrt{b_1}$  – průměr z 3. mocnin z-skórů

[SKEW()]

[mean(scale(x)^3)]

$$\sqrt{b_1} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^3$$

- ▶ pro symetrický histogram  $\sqrt{b_1}$  blízké nule
- ▶ doprava protažený histogram pro  $\sqrt{b_1} \gg 0$
- ▶ doleva protažený histogram pro  $\sqrt{b_1} \ll 0$

## charakteristiky tvaru: šikmost

- ▶ invariantní vůči posunutí i změně měřítka:

$$\gamma(a + x) = \gamma(x)$$

$$\gamma(b \cdot x) = \gamma(x) \quad b > 0$$

- ▶ **šikmost**  $\sqrt{b_1}$  – průměr z 3. mocnin z-skóru

[SKEW()]

[mean(scale(x)^3)]

$$\sqrt{b_1} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^3$$

- ▶ pro symetrický histogram  $\sqrt{b_1}$  blízké nule
- ▶ doprava protažený histogram pro  $\sqrt{b_1} \gg 0$
- ▶ doleva protažený histogram pro  $\sqrt{b_1} \ll 0$

## charakteristiky tvaru: šikmost

- ▶ invariantní vůči posunutí i změně měřítka:

$$\gamma(a + x) = \gamma(x)$$

$$\gamma(b \cdot x) = \gamma(x) \quad b > 0$$

- ▶ **šikmost**  $\sqrt{b_1}$  – průměr z 3. mocnin z-skórů

[SKEW()]

[mean(scale(x)^3)]

$$\sqrt{b_1} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^3$$

- ▶ pro symetrický histogram  $\sqrt{b_1}$  blízké nule
- ▶ doprava protažený histogram pro  $\sqrt{b_1} \gg 0$
- ▶ doleva protažený histogram pro  $\sqrt{b_1} \ll 0$



## charakteristiky tvaru: šikmost

- ▶ invariantní vůči posunutí i změně měřítka:

$$\gamma(a + x) = \gamma(x)$$

$$\gamma(b \cdot x) = \gamma(x) \quad b > 0$$

- ▶ **šikmost**  $\sqrt{b_1}$  – průměr z 3. mocnin z-skóru

[SKEW()]

[mean(scale(x)^3)]

$$\sqrt{b_1} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^3$$

- ▶ pro symetrický histogram  $\sqrt{b_1}$  blízké nule
- ▶ doprava protažený histogram pro  $\sqrt{b_1} \gg 0$
- ▶ doleva protažený histogram pro  $\sqrt{b_1} \ll 0$

## charakteristiky tvaru: šikmost

- ▶ invariantní vůči posunutí i změně měřítka:

$$\gamma(a + x) = \gamma(x)$$

$$\gamma(b \cdot x) = \gamma(x) \quad b > 0$$

- ▶ **šikmost**  $\sqrt{b_1}$  – průměr z 3. mocnin z-skóru

[SKEW()]

[mean(scale(x)^3)]

$$\sqrt{b_1} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^3$$

- ▶ pro symetrický histogram  $\sqrt{b_1}$  blízké nule
- ▶ doprava protažený histogram pro  $\sqrt{b_1} \gg 0$
- ▶ doleva protažený histogram pro  $\sqrt{b_1} \ll 0$

## charakteristiky tvaru: špičatost

- ▶ **špičatost**  $b_2$  – **průměr** ze 4. mocnin z-skórů  
(někdy se odečítá 3) **[KURT()]** **[mean(scale(x)^4)]**

$$b_2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^4$$

- ▶ někdy se počítají odhady populační šikmosti a špičatosti jinak  
(Excel:  $s_x$  jinak, Fisherovo  $g_1, g_2$  – pro zajímavost)

$$g_1 = \frac{\sqrt{n(n-1)}}{n-2} \sqrt{b_1}, \quad g_2 = \frac{(n+1)(n-1)}{(n-2)(n-3)} \left( b_2 - \frac{3(n-1)}{n+1} \right)$$

- ▶ šikmost a špičatost slouží k hodnocení, zda lze předpokládat *normální rozdělení* (bude zavedeno později)

## charakteristiky tvaru: špičatost

- ▶ **špičatost**  $b_2$  – **průměr** ze 4. mocnin z-skórů (někdy se odečítá 3) **[KURT()]** **[mean(scale(x)^4)]**

$$b_2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^4$$

- ▶ někdy se počítají odhady populační šikmosti a špičatosti jinak (Excel:  $s_x$  jinak, Fisherovo  $g_1, g_2$  – pro zajímavost)

$$g_1 = \frac{\sqrt{n(n-1)}}{n-2} \sqrt{b_1}, \quad g_2 = \frac{(n+1)(n-1)}{(n-2)(n-3)} \left( b_2 - \frac{3(n-1)}{n+1} \right)$$

- ▶ šikmost a špičatost slouží k hodnocení, zda lze předpokládat *normální rozdělení* (bude zavedeno později)

## charakteristiky tvaru: špičatost

- ▶ **špičatost**  $b_2$  – **průměr** ze 4. mocnin z-skórů (někdy se odečítá 3) **[KURT()]** [mean(scale(x)^4)]

$$b_2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right)^4$$

- ▶ někdy se počítají odhady populační šikmosti a špičatosti jinak (Excel:  $s_x$  jinak, Fisherovo  $g_1, g_2$  – pro zajímavost)

$$g_1 = \frac{\sqrt{n(n-1)}}{n-2} \sqrt{b_1}, \quad g_2 = \frac{(n+1)(n-1)}{(n-2)(n-3)} \left( b_2 - \frac{3(n-1)}{n+1} \right)$$

- ▶ šikmost a špičatost slouží k hodnocení, zda lze předpokládat *normální rozdělení* (bude zavedeno později)

## přehled závislostí

- ▶ abychom mohli vyšetřovat závislost, musíme na jedné statistické jednotce měřit aspoň dva znaky
- ▶ postupy (i grafické) závisí na měřících obou znaků
  - ▶ kvalitativní – kvalitativní  
(vzdělání – pracovní zařazení)
  - ▶ kvalitativní – kvantitativní  
(vzdělání – roční příjem)
  - ▶ kvantitativní – kvantitativní  
(věk – roční příjem)
- ▶ zatím popisné charakteristiky a grafy, prokazování závislosti později

## přehled závislostí

- ▶ abychom mohli vyšetřovat závislost, musíme na jedné statistické jednotce měřit aspoň dva znaky
- ▶ postupy (i grafické) závisí na měřících obou znaků
  - ▶ kvalitativní – kvalitativní  
(vzdělání – pracovní zařazení)
  - ▶ kvalitativní – kvantitativní  
(vzdělání – roční příjem)
  - ▶ kvantitativní – kvantitativní  
(věk – roční příjem)
- ▶ zatím popisné charakteristiky a grafy, prokazování závislosti později

## přehled závislostí

- ▶ abychom mohli vyšetřovat závislost, musíme na jedné statistické jednotce měřit aspoň dva znaky
- ▶ postupy (i grafické) závisí na měřících obou znaků
  - ▶ kvalitativní – kvalitativní  
(vzdělání – pracovní zařazení)
  - ▶ kvalitativní – kvantitativní  
(vzdělání – roční příjem)
  - ▶ kvantitativní – kvantitativní  
(věk – roční příjem)
- ▶ zatím popisné charakteristiky a grafy, prokazování závislosti později



## přehled závislostí

- ▶ abychom mohli vyšetřovat závislost, musíme na jedné statistické jednotce měřit aspoň dva znaky
- ▶ postupy (i grafické) závisí na měřících obou znaků
  - ▶ kvalitativní – kvalitativní  
(vzdělání – pracovní zařazení)
  - ▶ kvalitativní – kvantitativní  
(vzdělání – roční příjem)
  - ▶ kvantitativní – kvantitativní  
(věk – roční příjem)
- ▶ zatím popisné charakteristiky a grafy, prokazování závislosti později

## přehled závislostí

- ▶ abychom mohli vyšetřovat závislost, musíme na jedné statistické jednotce měřit aspoň dva znaky
- ▶ postupy (i grafické) závisí na měřících obou znaků
  - ▶ kvalitativní – kvalitativní  
(vzdělání – pracovní zařazení)
  - ▶ kvalitativní – kvantitativní  
(vzdělání – roční příjem)
  - ▶ kvantitativní – kvantitativní  
(věk – roční příjem)
- ▶ zatím popisné charakteristiky a grafy, prokazování závislosti později

## přehled závislostí

- ▶ abychom mohli vyšetřovat závislost, musíme na jedné statistické jednotce měřit aspoň dva znaky
- ▶ postupy (i grafické) závisí na měřících obou znaků
  - ▶ kvalitativní – kvalitativní  
(vzdělání – pracovní zařazení)
  - ▶ kvalitativní – kvantitativní  
(vzdělání – roční příjem)
  - ▶ kvantitativní – kvantitativní  
(věk – roční příjem)
- ▶ zatím popisné charakteristiky a grafy, prokazování závislosti později

## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřítku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$  (sdružené četnosti)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ , **čtyřpolní tabulka** [fourfold table]

## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřítku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$   
(**sdružené četnosti**)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  
 $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ ,  
**čtyřpolní tabulka** [fourfold table]

## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřtku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$   
(**sdružené četnosti**)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  
 $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ ,  
**čtyřpolní tabulka** [fourfold table]

## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřítku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$   
(**sdružené četnosti**)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  
 $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ ,  
**čtyřpolní tabulka** [fourfold table]

## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřítku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$   
(**sdružené četnosti**)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  
 $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ ,  
**čtyřpolní tabulka** [fourfold table]



## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřtku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$   
(**sdružené četnosti**)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  
 $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ ,  
**čtyřpolní tabulka** [fourfold table]

## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřtku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$   
(**sdružené četnosti**)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  
 $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ ,  
**čtyřpolní tabulka** [fourfold table]

## kvalitativní – kvalitativní

- ▶ kvalitativní data – znak v nominálním (ordinálním) měřítku
- ▶ hodnoty vyjadřujeme pomocí četností
- ▶ dva znaky – četnosti možných **dvojic hodnot**  $n_{ij}$   
(**sdružené četnosti**)
- ▶ zapisujeme do **kontingenční tabulky** [contingency table]  
 $[table(x,y)]$  nebo  $[xtabs(\sim x+y)]$
- ▶ doplňujeme **marginální četnosti** [marginal frequencies]
  - ▶ součty po řádcích a po sloupcích
  - ▶ četnosti jednotlivých hodnot každého ze znaků zvlášť
- ▶ oba znaky nula-jedničkové – kontingenční tabulka  $2 \times 2$ ,  
**čtyřpolní tabulka** [fourfold table]

## příklad – vzdělání matek

(pozor na orientaci grafu!)

vzdělání	porodnice		celkem
	Praha	venkov	
základní	23	11	34
střední	30	17	47
VŠ	17	1	18
celkem	70	29	99

vzdělání	porodnice		celkem
	Praha	venkov	
základní	32,9 %	37,9 %	34,3 %
střední	42,8 %	58,6 %	47,5 %
VŠ	24,3 %	3,5 %	18,2 %
celkem	100 %	100 %	100 %

## příklad – vzdělání matek

(pozor na orientaci grafu!)

vzdělání	porodnice		celkem
	Praha	venkov	
základní	23	11	34
střední	30	17	47
VŠ	17	1	18
celkem	70	29	99

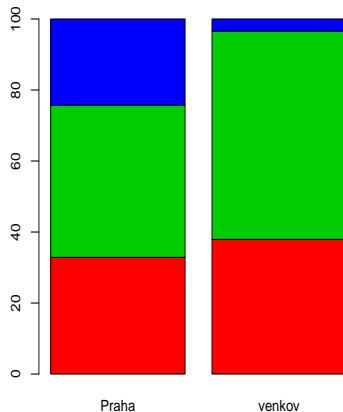
vzdělání	porodnice		celkem
	Praha	venkov	
základní	32,9 %	37,9 %	34,3 %
střední	42,8 %	58,6 %	47,5 %
VŠ	24,3 %	3,5 %	18,2 %
celkem	100 %	100 %	100 %

## příklad – vzdělání matek

(pozor na orientaci grafu!)

vzdělání	porodnice		celkem
	Praha	venkov	
základní	23	11	34
střední	30	17	47
VŠ	17	1	18
celkem	70	29	99

vzdělání	porodnice		celkem
	Praha	venkov	
základní	32,9 %	37,9 %	34,3 %
střední	42,8 %	58,6 %	47,5 %
VŠ	24,3 %	3,5 %	18,2 %
celkem	100 %	100 %	100 %



## příklad – vzdělání matek

(pozor na orientaci)

vzdělání	porodnice		celkem
	Praha	venkov	
základní	23	11	34
střední	30	17	47
VŠ	17	1	18
celkem	70	29	99

vzdělání	porodnice		celkem
	Praha	venkov	
základní	67,6 %	32,4 %	100 %
střední	63,8 %	36,2 %	100 %
VŠ	94,4 %	6,6 %	100 %
celkem	70,7 %	29,3 %	100 %

## příklad – vzdělání matek

(pozor na orientaci)

vzdělání	porodnice		celkem
	Praha	venkov	
základní	23	11	34
střední	30	17	47
VŠ	17	1	18
celkem	70	29	99

vzdělání	porodnice		celkem
	Praha	venkov	
základní	67,6 %	32,4 %	100 %
střední	63,8 %	36,2 %	100 %
VŠ	94,4 %	6,6 %	100 %
celkem	70,7 %	29,3 %	100 %

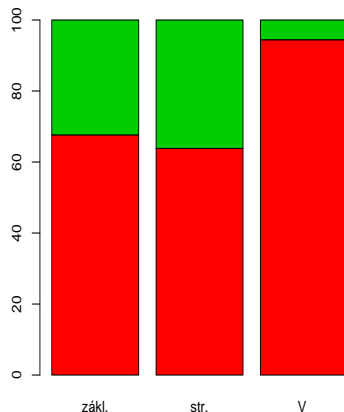


## příklad – vzdělání matek

(pozor na orientaci)

vzdělání	porodnice		celkem
	Praha	venkov	
základní	23	11	34
střední	30	17	47
VŠ	17	1	18
celkem	70	29	99

vzdělání	porodnice		celkem
	Praha	venkov	
základní	67,6 %	32,4 %	100 %
střední	63,8 %	36,2 %	100 %
VŠ	94,4 %	6,6 %	100 %
celkem	70,7 %	29,3 %	100 %



## kvalitativní – kvantitativní

- ▶ podle kvalitativní proměnné rozdělíme hodnoty kvantitativní proměnné do dílčích souborů
- ▶ porovnáme charakteristiky dílčích souborů (zejména charakteristiky polohy) mezi sebou, pokud se hodně liší, svědčí to pro závislost
- ▶ celkový průměr = vážený průměr dílčích souborů
- ▶ celkový rozptyl  $\doteq$  vážený průměr rozptylů + vážený rozptyl průměrů (přesně jen pro populační rozptyly s  $n$  ve jmenovateli)
- ▶ snáze jako **rozklad součtu čtverců**

## kvalitativní – kvantitativní

- ▶ podle kvalitativní proměnné rozdělíme hodnoty kvantitativní proměnné do dílčích souborů
- ▶ porovnáme charakteristiky dílčích souborů (zejména charakteristiky polohy) mezi sebou, pokud se hodně liší, svědčí to pro závislost
- ▶ celkový průměr = vážený průměr dílčích souborů
- ▶ celkový rozptyl  $\doteq$  vážený průměr rozptylů + vážený rozptyl průměrů (přesně jen pro populační rozptyly s  $n$  ve jmenovateli)
- ▶ snáze jako **rozklad součtu čtverců**

## kvalitativní – kvantitativní

- ▶ podle kvalitativní proměnné rozdělíme hodnoty kvantitativní proměnné do dílčích souborů
- ▶ porovnáme charakteristiky dílčích souborů (zejména charakteristiky polohy) mezi sebou, pokud se hodně liší, svědčí to pro závislost
- ▶ celkový průměr = vážený průměr dílčích souborů
- ▶ celkový rozptyl  $\doteq$  vážený průměr rozptylů + vážený rozptyl průměrů (přesně jen pro populační rozptyly s  $n$  ve jmenovateli)
- ▶ snáze jako **rozklad součtu čtverců**

## kvalitativní – kvantitativní

- ▶ podle kvalitativní proměnné rozdělíme hodnoty kvantitativní proměnné do dílčích souborů
- ▶ porovnáme charakteristiky dílčích souborů (zejména charakteristiky polohy) mezi sebou, pokud se hodně liší, svědčí to pro závislost
- ▶ celkový průměr = vážený průměr dílčích souborů
- ▶ celkový rozptyl  $\doteq$  vážený průměr rozptylů + vážený rozptyl průměrů (přesně jen pro populační rozptyly s  $n$  ve jmenovateli)
- ▶ snáze jako rozklad součtu čtverců

## kvalitativní – kvantitativní

- ▶ podle kvalitativní proměnné rozdělíme hodnoty kvantitativní proměnné do dílčích souborů
- ▶ porovnáme charakteristiky dílčích souborů (zejména charakteristiky polohy) mezi sebou, pokud se hodně liší, svědčí to pro závislost
- ▶ celkový průměr = vážený průměr dílčích souborů
- ▶ celkový rozptyl  $\doteq$  vážený průměr rozptylů + vážený rozptyl průměrů (přesně jen pro populační rozptyly s  $n$  ve jmenovateli)
- ▶ snáze jako **rozklad součtu čtverců**

## příklad: platy u tří skupin zaměstnanců

skup.	příjem	$n_j$	$\bar{x}_j$	$s_j$	$s_j^2$
žlutí	200 150	2	175,00	35,4	1250,0
modří	80 70 60 60	4	67,50	9,6	91,7
černí	20 20 18 18 15 15 10 10	8	15,75	4,0	16,2
celkem	746	14	53,29	57,7	3334,4

$$\bar{x} = \frac{2 \cdot 175,0 + 4 \cdot 67,50 + 8 \cdot 15,75}{2 + 4 + 8} = 53,29$$

$$s^2 = 3334,4 > \frac{2 \cdot 1250,0 + 4 \cdot 91,7 + 8 \cdot 16,2}{2 + 4 + 8} = 214,0$$

- ▶ nevážený (nesmyslný) průměr by byl 86,08!
- ▶ rozptyl celkem je mnohem větší, než jsou rozptyly ve skupinách
- ▶ příčina: nestejně průměry

## příklad: platy u tří skupin zaměstnanců

skup.	příjem	$n_j$	$\bar{x}_j$	$s_j$	$s_j^2$
žlutí	200 150	2	175,00	35,4	1250,0
modří	80 70 60 60	4	67,50	9,6	91,7
černí	20 20 18 18 15 15 10 10	8	15,75	4,0	16,2
celkem	746	14	53,29	57,7	3334,4

$$\bar{x} = \frac{2 \cdot 175,0 + 4 \cdot 67,50 + 8 \cdot 15,75}{2 + 4 + 8} = 53,29$$

$$s^2 = 3334,4 > \frac{2 \cdot 1250,0 + 4 \cdot 91,7 + 8 \cdot 16,2}{2 + 4 + 8} = 214,0$$

- ▶ nevážený (nesmyslný) průměr by byl 86,08!
- ▶ rozptyl celkem je mnohem větší, než jsou rozptyly ve skupinách
- ▶ příčina: nesteréžné průměry



## příklad: platy u tří skupin zaměstnanců

skup.	příjem	$n_j$	$\bar{x}_j$	$s_j$	$s_j^2$
žlutí	200 150	2	175,00	35,4	1250,0
modří	80 70 60 60	4	67,50	9,6	91,7
černí	20 20 18 18 15 15 10 10	8	15,75	4,0	16,2
celkem	746	14	53,29	57,7	3334,4

$$\bar{x} = \frac{2 \cdot 175,0 + 4 \cdot 67,50 + 8 \cdot 15,75}{2 + 4 + 8} = 53,29$$

$$s^2 = 3334,4 > \frac{2 \cdot 1250,0 + 4 \cdot 91,7 + 8 \cdot 16,2}{2 + 4 + 8} = 214,0$$

- ▶ nevážený (nesmyslný) průměr by byl 86,08!
- ▶ rozptyl celkem je mnohem větší, než jsou rozptyly ve skupinách
- ▶ příčina: nesteržné průměry

## příklad: platy u tří skupin zaměstnanců

skup.	příjem	$n_j$	$\bar{x}_j$	$s_j$	$s_j^2$
žlutí	200 150	2	175,00	35,4	1250,0
modří	80 70 60 60	4	67,50	9,6	91,7
černí	20 20 18 18 15 15 10 10	8	15,75	4,0	16,2
celkem	746	14	53,29	57,7	3334,4

$$\bar{x} = \frac{2 \cdot 175,0 + 4 \cdot 67,50 + 8 \cdot 15,75}{2 + 4 + 8} = 53,29$$

$$s^2 = 3334,4 > \frac{2 \cdot 1250,0 + 4 \cdot 91,7 + 8 \cdot 16,2}{2 + 4 + 8} = 214,0$$

- ▶ nevážený (nesmyslný) průměr by byl 86,08!
- ▶ rozptyl celkem je mnohem větší, než jsou rozptyly ve skupinách
- ▶ příčina: nesteržné průměry

## příklad: platy u tří skupin zaměstnanců

skup.	příjem	$n_j$	$\bar{x}_j$	$s_j$	$s_j^2$
žlutí	200 150	2	175,00	35,4	1250,0
modří	80 70 60 60	4	67,50	9,6	91,7
černí	20 20 18 18 15 15 10 10	8	15,75	4,0	16,2
celkem	746	14	53,29	57,7	3334,4

$$\bar{x} = \frac{2 \cdot 175,0 + 4 \cdot 67,50 + 8 \cdot 15,75}{2 + 4 + 8} = 53,29$$

$$s^2 = 3334,4 > \frac{2 \cdot 1250,0 + 4 \cdot 91,7 + 8 \cdot 16,2}{2 + 4 + 8} = 214,0$$

- ▶ nevážený (nesmyslný) průměr by byl 86,08!
- ▶ rozptyl celkem je mnohem větší, než jsou rozptyly ve skupinách
- ▶ příčina: nesteréžné průměry

## příklad: platy u tří skupin zaměstnanců

skup.	příjem	$n_j$	$\bar{x}_j$	$s_j$	$s_j^2$
žlutí	200 150	2	175,00	35,4	1250,0
modří	80 70 60 60	4	67,50	9,6	91,7
černí	20 20 18 18 15 15 10 10	8	15,75	4,0	16,2
celkem	746	14	53,29	57,7	3334,4

$$\bar{x} = \frac{2 \cdot 175,0 + 4 \cdot 67,50 + 8 \cdot 15,75}{2 + 4 + 8} = 53,29$$

$$s^2 = 3334,4 > \frac{2 \cdot 1250,0 + 4 \cdot 91,7 + 8 \cdot 16,2}{2 + 4 + 8} = 214,0$$

- ▶ nevážený (nesmyslný) průměr by byl 86,08!
- ▶ rozptyl celkem je mnohem větší, než jsou rozptyly ve skupinách
- ▶ příčina: nesteréžné průměry

## rozklad součtu čtverců

- ▶ velikost kolísání **všech** platů (celková variabilita):

$$\begin{aligned} SST &= (200 - 53,29)^2 + (150 - 53,29)^2 + (80 - 53,29)^2 + \dots \\ &\quad + (10 - 53,29)^2 = 43\,346,86 \end{aligned}$$

- ▶ velikost kolísání **uvnitř** skupin:

$$\begin{aligned} SSE &= (200 - 175)^2 + (150 - 175)^2 + (80 - 67,5)^2 + \dots \\ &\quad + (10 - 15,75)^2 = 1\,638,5 \end{aligned}$$

- ▶ kolísání průměrů (**mezi** skupinami):

$$\begin{aligned} SSA &= 2 \cdot (175 - 53,29)^2 + 4 \cdot (67,5 - 53,29)^2 \\ &\quad + 8 \cdot (15,75 - 53,29)^2 = 41\,708,36 \end{aligned}$$

- ▶ kontrola:  $1\,638,5 + 41\,708,36 = 43\,346,86$

## rozklad součtu čtverců

- ▶ velikost kolísání **všech** platů (celková variabilita):

$$\begin{aligned} SST &= (200 - 53,29)^2 + (150 - 53,29)^2 + (80 - 53,29)^2 + \dots \\ &\quad + (10 - 53,29)^2 = 43\,346,86 \end{aligned}$$

- ▶ velikost kolísání **uvnitř** skupin:

$$\begin{aligned} SSE &= (200 - 175)^2 + (150 - 175)^2 + (80 - 67,5)^2 + \dots \\ &\quad + (10 - 15,75)^2 = 1\,638,5 \end{aligned}$$

- ▶ kolísání průměrů (**mezi** skupinami):

$$\begin{aligned} SSA &= 2 \cdot (175 - 53,29)^2 + 4 \cdot (67,5 - 53,29)^2 \\ &\quad + 8 \cdot (15,75 - 53,29)^2 = 41\,708,36 \end{aligned}$$

- ▶ kontrola:  $1\,638,5 + 41\,708,36 = 43\,346,86$

## rozklad součtu čtverců

- ▶ velikost kolísání **všech** platů (celková variabilita):

$$\begin{aligned} SST &= (200 - 53,29)^2 + (150 - 53,29)^2 + (80 - 53,29)^2 + \dots \\ &\quad + (10 - 53,29)^2 = 43\,346,86 \end{aligned}$$

- ▶ velikost kolísání **uvnitř** skupin:

$$\begin{aligned} SSE &= (200 - 175)^2 + (150 - 175)^2 + (80 - 67,5)^2 + \dots \\ &\quad + (10 - 15,75)^2 = 1\,638,5 \end{aligned}$$

- ▶ kolísání průměrů (**mezi** skupinami):

$$\begin{aligned} SSA &= 2 \cdot (175 - 53,29)^2 + 4 \cdot (67,5 - 53,29)^2 \\ &\quad + 8 \cdot (15,75 - 53,29)^2 = 41\,708,36 \end{aligned}$$

- ▶ kontrola:  $1\,638,5 + 41\,708,36 = 43\,346,86$

## rozklad součtu čtverců

- ▶ velikost kolísání **všech** platů (celková variabilita):

$$\begin{aligned} SST &= (200 - 53,29)^2 + (150 - 53,29)^2 + (80 - 53,29)^2 + \dots \\ &\quad + (10 - 53,29)^2 = 43\,346,86 \end{aligned}$$

- ▶ velikost kolísání **uvnitř** skupin:

$$\begin{aligned} SSE &= (200 - 175)^2 + (150 - 175)^2 + (80 - 67,5)^2 + \dots \\ &\quad + (10 - 15,75)^2 = 1\,638,5 \end{aligned}$$

- ▶ kolísání průměrů (**mezi** skupinami):

$$\begin{aligned} SSA &= 2 \cdot (175 - 53,29)^2 + 4 \cdot (67,5 - 53,29)^2 \\ &\quad + 8 \cdot (15,75 - 53,29)^2 = 41\,708,36 \end{aligned}$$

- ▶ kontrola:  $1\,638,5 + 41\,708,36 = 43\,346,86$



## rozklad součtu čtverců obecně

- ▶  $x_{ij}$   $j$ -tá hodnota v  $i$ -té skupině (plat  $j$ -té osoby v  $i$ -té skupině)
- ▶  $n_i$  počet hodnot v  $i$ -té skupině,  $k$  počet skupin
- ▶  $\bar{x}_{i\bullet}$  průměr v  $i$ -té skupině (průměrný plat v  $i$ -té skupině)
- ▶  $\bar{x}_{\bullet\bullet}$  celkový průměr (průměr všech platů)

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{\bullet\bullet})^2 \\ &= \sum_{i=1}^k n_i (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\bullet})^2 \\ &= SSA + SSE \end{aligned}$$

## rozklad součtu čtverců obecně

- ▶  $x_{ij}$   $j$ -tá hodnota v  $i$ -té skupině (plat  $j$ -té osoby v  $i$ -té skupině)
- ▶  $n_i$  počet hodnot v  $i$ -té skupině,  $k$  počet skupin
- ▶  $\bar{x}_{i\bullet}$  průměr v  $i$ -té skupině (průměrný plat v  $i$ -té skupině)
- ▶  $\bar{x}_{\bullet\bullet}$  celkový průměr (průměr všech platů)

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{\bullet\bullet})^2 \\ &= \sum_{i=1}^k n_i (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\bullet})^2 \\ &= SSA + SSE \end{aligned}$$

## rozklad součtu čtverců obecně

- ▶  $x_{ij}$   $j$ -tá hodnota v  $i$ -té skupině (plat  $j$ -té osoby v  $i$ -té skupině)
- ▶  $n_i$  počet hodnot v  $i$ -té skupině,  $k$  počet skupin
- ▶  $\bar{x}_{i\bullet}$  průměr v  $i$ -té skupině (průměrný plat v  $i$ -té skupině)
- ▶  $\bar{x}_{\bullet\bullet}$  celkový průměr (průměr všech platů)

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{\bullet\bullet})^2 \\ &= \sum_{i=1}^k n_i (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\bullet})^2 \\ &= SSA + SSE \end{aligned}$$

## rozklad součtu čtverců obecně

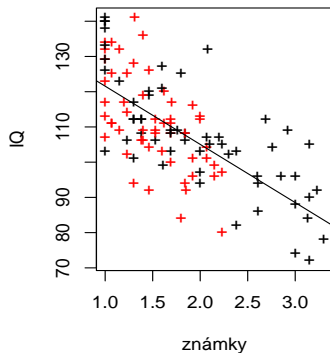
- ▶  $x_{ij}$   $j$ -tá hodnota v  $i$ -té skupině (plat  $j$ -té osoby v  $i$ -té skupině)
- ▶  $n_i$  počet hodnot v  $i$ -té skupině,  $k$  počet skupin
- ▶  $\bar{x}_{i\bullet}$  průměr v  $i$ -té skupině (průměrný plat v  $i$ -té skupině)
- ▶  $\bar{x}_{\bullet\bullet}$  celkový průměr (průměr všech platů)

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{\bullet\bullet})^2 \\ &= \sum_{i=1}^k n_i (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\bullet})^2 \\ &= SSA + SSE \end{aligned}$$

## kvantitativní – kvantitativní

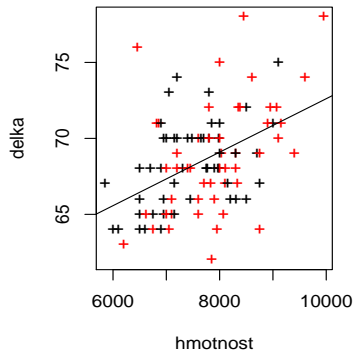
```
plot(iq~zn7,data=lq,col=1+divka,pch="+")
```

záporná korelace



$$r = -0,69$$

kladná korelace



$$r = 0,45$$

## popis závislosti spojených veličin

- ▶ (výběrová) **kovariance** [covariance] [cov(vek.o,vek.m)]

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ zřejmě je  $s_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = s_x^2$ ,  $s_{yy} = s_y^2$

- ▶ (Pearsonův, momentový) **korelační koeficient**  
[(Pearson, product-moment) correlation coefficient]

- ▶ lze zapsat pomocí z-skórů [cor(vek.o,vek.m)]

$$r = \frac{s_{xy}}{s_x s_y} = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \cdot \frac{y_i - \bar{y}}{s_y} \right)$$

## popis závislosti spojitých veličin

- ▶ (výběrová) **kovariance** [covariance] [cov(vek.o,vek.m)]

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ zřejmě je  $s_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = s_x^2$ ,  $s_{yy} = s_y^2$

- ▶ (Pearsonův, momentový) **korelační koeficient**  
 [(Pearson, product-moment) correlation coefficient]

- ▶ lze zapsat pomocí z-skórů [cor(vek.o,vek.m)]

$$r = \frac{s_{xy}}{s_x s_y} = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \cdot \frac{y_i - \bar{y}}{s_y} \right)$$

## popis závislosti spojitých veličin

- ▶ (výběrová) **kovariance** [covariance] [cov(vek.o,vek.m)]

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ zřejmě je  $s_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = s_x^2$ ,  $s_{yy} = s_y^2$

- ▶ (Pearsonův, momentový) **korelační koeficient**  
 [(Pearson, product-moment) correlation coefficient]

- ▶ lze zapsat pomocí z-skórů [cor(vek.o,vek.m)]

$$r = \frac{s_{xy}}{s_x s_y} = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \cdot \frac{y_i - \bar{y}}{s_y} \right)$$



## popis závislosti spojitých veličin

- ▶ (výběrová) **kovariance** [covariance] [cov(vek.o,vek.m)]

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- ▶ zřejmě je  $s_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = s_x^2$ ,  $s_{yy} = s_y^2$

- ▶ (Pearsonův, momentový) **korelační koeficient**  
[(Pearson, product-moment) correlation coefficient]

- ▶ lze zapsat pomocí z-skórů [cor(vek.o,vek.m)]

$$r = \frac{s_{xy}}{s_x s_y} = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \cdot \frac{y_i - \bar{y}}{s_y} \right)$$

## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?

## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?

## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?

## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?

## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?

## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?

## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?



## příklad: hmotnost a délka dětí (24. týden věku)

- ▶ délka [cm]:  $\bar{x} = 68,5$   $s_x = 3,28$
- ▶ hmotnost [g]:  $\bar{y} = 7690$ ,  $s_y = 845$
- ▶ kovariance [cm · g]:  $s_{xy} = 1257$
- ▶ korelační koeficient:  $r = \frac{1257}{3,28 \cdot 845} = 0,45$
- ▶ hmotnost [kg]:  $\bar{y} = 7,69$   $s_y = 0,845$
- ▶ kovariance [cm · kg]:  $s_{xy} = 1,257$
- ▶ korelační koeficient:  $r = \frac{1,257}{3,28 \cdot 0,845} = 0,45$
- ▶ které charakteristiky závisí na použitém měřítku?

# vlastnosti Pearsonova korelačního koeficient

- ▶ vypovídá o směru závislosti
- ▶ při  $r < 0$  s rostoucím  $x$  v průměru  $y$  klesá (např. IQ a známky)
- ▶ při  $r > 0$  s rostoucím  $x$  v průměru  $y$  roste (např. váha a výška)
- ▶ platí  $-1 \leq r \leq 1$
- ▶  $|r| = 1$  jedině tehdy, když body  $[x; y]$  leží na přímce
- ▶ vzájemné nezávislosti  $x, y$  odpovídají  $r$  blízka nule (upřesníme!)
- ▶ nemusí zachytit křivočarou (nelineární) závislost

## vlastnosti Pearsonova korelačního koeficient

- ▶ vypovídá o směru závislosti
- ▶ při  $r < 0$  s rostoucím  $x$  v průměru  $y$  klesá (např. IQ a známky)
- ▶ při  $r > 0$  s rostoucím  $x$  v průměru  $y$  roste (např. váha a výška)
- ▶ platí  $-1 \leq r \leq 1$
- ▶  $|r| = 1$  jedině tehdy, když body  $[x; y]$  leží na přímce
- ▶ vzájemné nezávislosti  $x, y$  odpovídají  $r$  blízka nule (upřesníme!)
- ▶ nemusí zachytit křivočarou (nelineární) závislost

## vlastnosti Pearsonova korelačního koeficient

- ▶ vypovídá o směru závislosti
- ▶ při  $r < 0$  s rostoucím  $x$  v průměru  $y$  klesá (např. IQ a známky)
- ▶ při  $r > 0$  s rostoucím  $x$  v průměru  $y$  roste (např. váha a výška)
- ▶ platí  $-1 \leq r \leq 1$
- ▶  $|r| = 1$  jedině tehdy, když body  $[x; y]$  leží na přímce
- ▶ vzájemné nezávislosti  $x, y$  odpovídají  $r$  blízka nule (upřesníme!)
- ▶ nemusí zachytit křivočarou (nelineární) závislost

## vlastnosti Pearsonova korelačního koeficient

- ▶ vypovídá o směru závislosti
- ▶ při  $r < 0$  s rostoucím  $x$  v průměru  $y$  klesá (např. IQ a známky)
- ▶ při  $r > 0$  s rostoucím  $x$  v průměru  $y$  roste (např. váha a výška)
- ▶ platí  $-1 \leq r \leq 1$
- ▶  $|r| = 1$  jedině tehdy, když body  $[x; y]$  leží na přímce
- ▶ vzájemné nezávislosti  $x, y$  odpovídají  $r$  blízka nule (upřesníme!)
- ▶ nemusí zachytit křivočarou (nelineární) závislost

## vlastnosti Pearsonova korelačního koeficient

- ▶ vypovídá o směru závislosti
- ▶ při  $r < 0$  s rostoucím  $x$  v průměru  $y$  klesá (např. IQ a známky)
- ▶ při  $r > 0$  s rostoucím  $x$  v průměru  $y$  roste (např. váha a výška)
- ▶ platí  $-1 \leq r \leq 1$
- ▶  $|r| = 1$  jedině tehdy, když body  $[x; y]$  leží na přímce
- ▶ vzájemné nezávislosti  $x, y$  odpovídají  $r$  blízka nule (upřesníme!)
- ▶ nemusí zachytit křivočarou (nelineární) závislost

## vlastnosti Pearsonova korelačního koeficient

- ▶ vypovídá o směru závislosti
- ▶ při  $r < 0$  s rostoucím  $x$  v průměru  $y$  klesá (např. IQ a známky)
- ▶ při  $r > 0$  s rostoucím  $x$  v průměru  $y$  roste (např. váha a výška)
- ▶ platí  $-1 \leq r \leq 1$
- ▶  $|r| = 1$  jedině tehdy, když body  $[x; y]$  leží na přímce
- ▶ vzájemné nezávislosti  $x, y$  odpovídají  $r$  blízka nule (upřesníme!)
- ▶ nemusí zachytit křivočarou (nelineární) závislost

## vlastnosti Pearsonova korelačního koeficient

- ▶ vypovídá o směru závislosti
- ▶ při  $r < 0$  s rostoucím  $x$  v průměru  $y$  klesá (např. IQ a známky)
- ▶ při  $r > 0$  s rostoucím  $x$  v průměru  $y$  roste (např. váha a výška)
- ▶ platí  $-1 \leq r \leq 1$
- ▶  $|r| = 1$  jedině tehdy, když body  $[x; y]$  leží na přímce
- ▶ vzájemné nezávislosti  $x, y$  odpovídají  $r$  blízka nule (upřesníme!)
- ▶ nemusí zachytit křivočarou (nelineární) závislost